

**Complete mitochondrial genome sequence of the polychaete  
annelid *Platynereis dumerilii***

Jeffrey L. Boore

Department of Biology, University of Michigan

Ann Arbor MI 48109-1048, USA

and

DOE Joint Genome Institute and Lawrence Berkeley National Laboratory

2800 Mitchell Drive

Walnut Creek, CA 95498 USA

Keywords: *Platynereis*, annelid, mitochondria, evolution, genome

Running title: *Platynereis* mtDNA

2,650 total words, 2 figures

Corresponding author:

Jeffrey L. Boore

DOE Joint Genome Institute and Lawrence Berkeley National Laboratory

2800 Mitchell Drive

Walnut Creek, CA 95498 USA

jlboore@lbl.gov

(925) 296-5691

FAX: (925) 296-5620

Complete mitochondrial genome sequences are now available for 126 metazoans (see Boore 1999; Mitochondrial Genomics link at <http://www.jgi.doe.gov>), but the taxonomic representation is highly biased. For example, 80 are from a single phylum, Chordata, and show little variation for many molecular features. Arthropoda is represented by 16 taxa, Mollusca by eight, and Echinodermata by five, with only 17 others from the remaining ~30 metazoan phyla. With few exceptions (see Wolstenholme 1992 and Boore 1999) these are circular DNA molecules, about 16 kb in size, and encode the same set of 37 genes. A variety of non-standard names are sometimes used for animal mitochondrial genes; see Boore (1999) for gene nomenclature and a table of synonyms.

Mitochondrial genome comparisons serve as a model of genome evolution. In this system, much smaller and simpler than that of the nucleus, are all of the same factors of genome evolution, where one may find tractable the changes in tRNA structure, base composition, genetic code, gene arrangement, etc. Further, patterns of mitochondrial gene rearrangements are an exceptionally reliable indicator of phylogenetic relationships (Smith et al. 1993; Boore et al. 1995; Boore, Lavrov, and Brown 1998; Boore and Brown 1998, 2000; Dowton 1999; Stechmann and Schlegel 1999; Kurabayashi and Ueshima 2000). To these ends, we are sampling further the variation among major animal groups in features of their mitochondrial genomes.

The phylum Annelida is traditionally divided into three classes: Oligochaeta (e.g., earthworms); Hirudinida (leeches); and Polychaeta (marine annelids). The complete mitochondrial genome of an oligochaete, *Lumbricus terrestris*, has been previously described (Boore and Brown 1995). More recently, an homologous portion (~50%) of the mtDNAs of the hirudinid *Helobdella robusta*, the polychaete *Platynereis dumerilii*, and the siboglinid annelid

*Galathealinum brachiosum* (previously considered to be of the phylum Pogonophora) have been compared (Boore and Brown 2000).

This earlier described portion of the *P. dumerilii* mtDNA spans from near the end of *rrnL* to the middle of *cob* in the direction of transcription (see fig. 1). An additional fragment of *rrnS* has since been amplified by PCR using primers 12SA and 12SB (see Palumbi 1996 for primer sequences) and its sequence determined. Oligonucleotides matching the obtained sequences were used in “long-PCR” (Barnes 1994) to generate fragments spanning *cob-rrnS* and *rrnS-rrnL*. PCR reaction conditions, purifications, sequencing reactions, gene identifications, and analysis were as previously described (Boore and Brown 2000). All nucleotides were determined on both strands by primer walking except for a few short regions meeting the following criteria: less than 200 nts in length; within 400 nts of the sequencing primer; and without any hint of ambiguity on the sequenced strand. Together, all obtained fragments represent the entire *P. dumerilii* mtDNA in generously overlapping segments.

The mitochondrial genome of the polychaete annelid *P. dumerilii* is 15,619 nts in size, with an A+T content of 64.1%. The oligochaete *L. terrestris*, the only other annelid whose complete mtDNA sequence has been determined (Boore and Brown 1995), has a somewhat smaller (14,998 nts) and slightly less A+T rich (61.6%) mt genome. These two mtDNAs have similar bias in the distributions of nucleotides between each of their two DNA strands: GC-skew ( $[G-C]/[G+C]$ ) is -0.14 for *P. dumerilii* and -0.18 for *L. terrestris* and TA-skew ( $[T-A]/[T+A]$ ) is 0.026 for *P. dumerilii* and 0.031 for *L. terrestris*. From this it is possible to speculate that these two mtDNAs experience similar mutational biases. The pattern of codon usage is also similar between *P. dumerilii* (not shown) and *L. terrestris* (Boore and Brown 1995, 2000) mtDNAs.

*P. dumerilii* mtDNA contains the 37 genes most commonly found in animal mitochondrial genomes, including *atp8*, which is not present in mtDNAs of nematodes (see Boore 1999 and references therein) or of the bivalve mollusk *Mytilus edulis* (Hoffmann, Boore, and Brown 1992). All genes are transcribed from the same DNA strand, as is the case for *L. terrestris*, but which is uncommon for animal mtDNAs. In all studied annelids (*P. dumerilii*, *L. terrestris*, *H. robusta*, and *G. brachiosum*), *atp8* and *atp6* are not adjacent and so, obviously, can not be translated from a bicistronic mRNA, as has been shown for some animals (Fearnley and Walker 1986).

All 37 genes are arranged identically in *P. dumerilii* and *L. terrestris* mtDNAs except for six tRNAs (fig. 1). To conceptually convert the mitochondrial gene arrangement of *P. dumerilii* into that of *L. terrestris* would require a switching of positions between two adjacent tRNAs, *trnA* and *trnS2(uga)*, movement of *trnC* and *trnM* to between *nad4* and *rrnS*, and local rearrangements of *trnY*, *trnG*, and *trnD*. Each of these last three tRNA genes is flanked by one or more non-coding regions in *P. dumerilii* mtDNA, as has been noted for recent translocations for other taxa (see Boore 1999). There are also differences in the positions of non-coding regions between these two mtDNAs.

Alternatives to ATG start codons are very common among metazoan mtDNAs, so it is unusual to find an ATG codon at the beginning of all 13 protein-encoding genes. This same atypical condition has been found also for all 13 protein-encoding genes of *L. terrestris*, as well as all of the sampled genes of *G. brachiosum* and all but one of *H. robusta* (Boore and Brown 2000). It is possible that Annelida may not share with many other animals (Wolstenholme 1992) the potential for using a variety of alternative initiation codons.

Eight of the protein gene sequences appear to end with a single T or a TA that is directly adjacent to the downstream gene. It is common for termination codons to be truncated (to T or TA) in metazoan mtDNAs; such codons are converted to complete (UAA) stop codons by polyadenylation after transcript processing (Ojala et al. 1981). However, five of these have a complete, in-frame stop codon that would require only a short overlap with the downstream gene, sometimes of only one or two nucleotides. It is not obvious how overlapping genes would be resolved to whole, gene-specific mRNAs, but this is apparently accomplished for the cases of overlapping tRNA genes (see below). It is possible that these overlapping, complete stop codons serve as “backups” to prevent translational readthrough in cases where the transcripts are not properly cleaved.

Animal mtDNAs almost universally contain the same set of 22 tRNA genes (see Boore 1999), the minimum set necessary for translation of the mitochondrially encoded proteins using the relaxed wobble rules for mitochondrial ribosomes. Twenty-two potential secondary structures similar to those of mt-tRNAs, with potential anticodon sequences identical to those of *L. terrestris* mtDNA, can be identified for *P. dumerilii*, and are proposed as tRNA genes (Boore and Brown 2000 and fig. 2). All have the potential to form a seven nt-pair acceptor stem, two with a single mismatch each, and a five nt-pair anticodon stem, six with a single mismatch each. The consistency in position of the mismatched nucleotides among the several tRNAs invites speculation that these mismatches may have a specific function.

The nucleotides preceding and following the anticodon are always a T and a purine, respectively. The nucleotide preceding the anticodon stem is an A in all but six of the tRNA genes. There are four nucleotides in the “extra” arm of all except five of the tRNAs, those for N, I, K, S1, and T; each of these has five nucleotides here. It seems noteworthy that these five

tRNAs are also the only ones that have other than T at the eighth nucleotide position, suggesting some interaction in structure or function of these correlated features.

There are three cases in which the sequences of adjacent tRNA genes overlap. For the pair *trnA-trnL2(uaa)* the overlap is of only the discriminator nucleotide, as discussed in Boore and Brown (2000). The other two pairs, *trnR-trnH* and *trnE-trnP*, each overlap by two nucleotides, so producing two full length tRNAs from these genes would require either posttranscriptional editing, independent transcriptional promoters, or differential transcript cleavage. Otherwise, assuming that all stop codons have been correctly inferred (see above), the only other overlapping gene pair is *nad4L-nad4*; this is common in animal mtDNAs, perhaps due to their translation from a bicistronic mRNA.

Inferring the precise ends of the ribosomal RNA transcripts from DNA sequence alone is not possible, but if it is assumed that these genes extend to the boundaries of the flanking genes, *P. dumerilii rrnS* is 790 nt with 63.1% A+T, very similar to *rrnS* of *L. terrestris*, which would be 785 nt in length with 59.6% A+T. These gene sequences have 59% identity between the two annelids. The sequences of the large subunit rRNA genes are slightly more variant; *P. dumerilii rrnL* would be 1172 nt (64.3% A+T), whereas *rrnL* of *L. terrestris* would be 1245 nt (64.9% A+T). These *rrnL* genes of the two annelids are 56% identical, although this value increases to 59% if the 52 nt of *L. terrestris rrnL* that align beyond the length of *P. dumerilii rrnL* are ignored; it may be possible that these “extensions” of *L. terrestris rrnL* may not be part of the transcript.

Despite their usually compact arrangement, the mtDNAs of all metazoans studied so far contain at least one large non-coding region. Many refer to this as the “control region” because, for a few species, it is known to contain elements that control replication and transcription (see

Shadel and Clayton 1997) and it is often assumed that this is generally its function. The largest non-coding (NC) region in *P. dumerilii* mtDNA is 1091 nt located between *trnG* and *trnY*. *L. terrestris* mtDNA has a much smaller analogous region, only 384 nt, and here it is between *trnR* and *trnH*.

The potential for a large RNA secondary structure had been previously noted within the large NC region of *L. terrestris* mtDNA (Boore and Brown 1995). One of the stems within this hypothetical structure can be formed in two different ways by alternative nucleotide pairings (fig. 2), reminiscent of regulatory signals in some prokaryotes (e.g. see Lewin 1987). A search of the large NC sequence of *P. dumerilii* for similar potential structure identifies a portion that also contains such an alternate pairing potential, although the predicted loops are of much smaller size. Although there is no experimental evidence to support that these structures actually form or serve any function, the conservation of similar potential structures between these two mtDNAs bolsters speculation that they may play some regulatory role.

*P. dumerilii* is only the second annelid, and the first from the class Polychaeta, for which a complete mtDNA sequence has been determined. This study reinforces some views of mtDNA evolution, such as how lineages long separated can have experienced very few gene rearrangements. Yet it offers some surprises, such as the common usage of standard (i.e. ATG) initiation codons appearing to characterize this phylum, conservation of unusual tRNA features suggesting that they serve some function, and the inference of a large, conserved secondary structure performing as a regulatory element. Future molecular experiments and mtDNA comparisons of other diverse metazoans will continue to illuminate the evolutionary processes of mitochondrial genomes.



## **Acknowledgments**

I am grateful to Daniel Sellos for *Platynereis dumerilii* DNA. The sequence of *P. dumerilii* mtDNA has been deposited in GenBank under accession number AF178678. This work was supported by DEB-9807100 from the National Science Foundation. Part of this work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

## Literature cited

- BARNES, W. M. 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from bacteriophage templates. *Proc. Natl. Acad. Sci. USA* **91**:2216-2220.
- BOORE, J. L. 1999. Animal mitochondrial genomes. *Nucl. Acids Res* **27**:1767-1780.
- BOORE, J. L., and W. M. BROWN. 1995. Complete DNA sequence of the mitochondrial genome of the annelid worm, *Lumbricus terrestris*. *Genetics* **141**:305-319.
- BOORE, J. L., and W. M. BROWN. 1998. Big trees from little genomes: Mitochondrial gene order as a phylogenetic tool. *Curr. Opin. Genet. Dev.* **8**:668-674.
- BOORE, J. L., and W. M. BROWN. 2000. Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: Sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. *Mol. Biol. Evol.* **17**(1):87-106.
- BOORE, J. L., T. M. COLLINS, D. STANTON, L. L. DAEHLER and W. M. BROWN. 1995. Deducing arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* **376**:163-165.
- BOORE, J. L., D. LAVROV and W. M. BROWN. 1998. Gene translocation links insects and crustaceans. *Nature* **392**:667-668.
- DOWTON, M. 1999. Relationships among the cyclostome braconid (Hymenoptera: Braconidae) subfamilies inferred from a mitochondrial tRNA gene rearrangement. *Mol. Phylogenet. Evol.* **11**:283-287.
- FEARNLEY, I. M., and J. E. WALKER. 1986. Two overlapping genes in bovine mitochondrial DNA encode membrane components of ATP synthase. *EMBO J.* **5**:2003-2008.
- HOFFMANN, R. J., J. L. BOORE and W. M. BROWN. 1992. A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. *Genetics* **131**:397-412.

- KURABAYASHI, A., and R. UESHIMA. 2000. Complete sequence of the mitochondrial DNA of the primitive opisthobranch *Pupa strigosa*: Systematic implications of the genome organization. *Mol. Biol. Evol.* **17**(2):266-277.
- LEWIN, B. 1987. GENES III. Wiley & Sons, Inc., New York. pp. 253-256.
- OJALA, D., J. MONTOYA and G. ATTARDI. 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**:470-474.
- PALUMBI, S. R. 1996. Nucleic acids II: The polymerase chain reaction. Pp. 205-247, in D. M. HILLIS, C. MORITZ and B. K. MABLE, eds. *Molecular Systematics*. Sinauer Associates, Sunderland, Massachusetts, USA.
- SHADEL, G. S., and D. A. CLAYTON. 1997. Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem.* **66**:409-435.
- SMITH, M. J., A. ARNDT, S. GORSKI and E. FAJBER. 1993. The phylogeny of echinoderm classes based on mitochondrial gene arrangements. *J. Mol. Evol.* **36**:545-554.
- STECHMANN, A, and M. SCHLEGEL. 1999. Analysis of the complete mitochondrial DNA sequence of the brachiopod *Terebratulina retusa* places Brachiopoda within the protostomes. *Proc. R. Soc. Lond., B, Biol. Sci.* **260**:2043-2052.
- WOLSTENHOLME, D. R. 1992. Animal mitochondrial DNA: structure and evolution, in *Mitochondrial Genomes*, (K. W. JEON and D. R. WOLSTENHOLME, eds.) *International Review of Cytology*. **141**:173-216.

## Figure Legends

FIG 1.—Gene map of the complete mitochondrial genome of the polychaete annelid *Platynereis dumerilii*. Genes are abbreviated as in the text, except that tRNA genes are designated by only the one-letter code for the corresponding amino acid, with L1/L2 and S1/S2 designating the tRNA pairs recognizing the leucine codons CUN/UUR (anticodons uag/uaa) and the serine codons AGN/UCN (anticodons ucu/uga), respectively. “NC” designates the largest non-coding region and numerals indicate the number of nucleotides in each of the other non-coding region larger than 30 nts. All genes are transcribed from the same strand and scaling is only approximate. An arrow marks the direction of transcription, but the origin(s) is not established. Underlining indicates the six tRNA genes whose relative positions differ from the arrangement found in the mitochondrial genome of the oligochaete annelid *Lumbricus terrestris* (Boore and Brown 1995); the arrangements of these genes in *L. terrestris* are shown on the periphery, as is the differing position of the largest non-coding region in *L. terrestris* mtDNA.

FIG 2.—Secondary structures found for portions of the mtDNA of *Platynereis dumerilii*. First is shown the eight mitochondrial tRNA genes that were not previously described in Boore and Brown (2000) folded into their potential secondary structures. Second is shown the similar potential secondary structures within the hypothetical transcripts of the largest non-coding regions of *P. dumerilii* and *L. terrestris* mtDNAs. Numerals in parentheses refer to the nucleotide positions beginning at the first nucleotide of *cox1*.



